CONVERGENCE ANALYSIS OF MARKOV CHAIN MONTE CARLO ESTIMATORS OF DISCRETE CHOICE MODELS IN TRANSPORTATION

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Chen Wang

May 2014

ABSTRACT

Although Bayes estimators are attractive for discrete choice models involving complex non-convex optimization and weak identification, researchers in transportation seem somewhat reluctant to adopt the Bayesian approach. A common argument against simulation-based Bayes estimators is that there are no general rules for assessing convergence. In this thesis, we study convergence of the Markov chain Monte Carlo (MCMC) estimator of logit and probit models, not only in marginal utility (preference) space but also in willingness-to-pay space. We use personal vehicle choice as case study, and we apply a series of convergence diagnostics. Because under regularity conditions the asymptotic distributions of frequentist and Bayes estimators coincide, we also compare the behavior of the posterior first and second moments with that of the point estimates of maximum (simulated) likelihood. When working in preference space, the Bayes estimators converge rather quickly. However, problems appear when analyzing convergence of willingness-to-pay measures that have not been discussed in previous literature. In particular, we observed that.

BIOGRAPHICAL SKETCH

Chen Wang was born in Xi’an, a historical city in China in 1989. During his childhood, he developed his great interest in how to construct buildings and then he performed excellent in related courses in the middle school and high school. To fulfill his dream, he went to Tsinghua University, the best university in China, and majored in civil engineering. In his senior year he change his mind that he want to solve the more and more serious traffic problems which caused by tremendous development of Chinese economics.

After graduation, he decided to go to Cornell University as an M.S. /Ph.D. student in the School of Civil & Environmental Engineering. He has concentrated on demand analysis of the vehicle market, developing and applying advanced econometric models. Specifically, in his master period, he paid more attention to Bayes estimators of discrete choice models as well as the related convergence issues. In 2014, he finished his M.S. degree in Cornell University and began to pursuing the doctoral degree in the same department with interest in gradient-free estimation of complex discrete choice models.

致我亲爱的爸爸和妈妈

For my dear father and mother

ACKNOWLEDGMENTS

I would like to thank my advisor Ricardo A. Daziano for his support during my research at Cornell. I always come up with some great research ideas after pleasing conversations with Ricardo. At the same time, I would like to thank all members in my special committee. Professor Huaizhu (Oliver) Gao was the temporary advisor of my first year's study and provide me with general introduction to transportation engineering. Professor Shanjun Li taught me abundant econometric knowledge when I was a rookie in economics. In addition, professor Li shares the economic way of thinking with me when it comes to practical problems.

I also appreciate all my officemates in Hollister Hall, especially Esther Chiew, Xi He, Yutaka Motoaki ,Bingyan Huang, Qing Zhao and Zhijie Dong. The frequent academic discussions brought me inspirations and solutions when I was stuck with my research. I am grateful for the fellow Ph.D. Xun Wang whom I learned a lot from.

Conducting endless research in Ithaca is not an easy task. I really appreciate my parents who always supported me emotionally when I was depressed. I would like to thank my fiancée Yuting Ji who brought color in my life. With her, I never feel lonely.

At last, thank all the people who gave me joy and help. Without you, I could never finish my degree here!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

WTP	Willingness-to-pay
MCMC	Markov chain Monte Carlo
EMRG	Energy and Materials Research Group
CAD2002	Canadian dollar in 2002
SGV	Standard Gasoline Vehicle
AFV	Alternative Fuel Vehicle
HEV	Hybrid Electric Vehicle
HFC	Hydrogen Fuel Cell Vehicle
S.D.	Standard Deviation
GHK	Geweke-Hajivassiliou-Keane
MLE	Maximum Simulated Likelihood Estimator
MSLE	Maximum Simulated Likelihood Estimator
Est.	Estimates
S.E.	Standard Error
MH	Metropolis-Hastings
CC	Capital Cost
FC	Fuel Cost
FA	Fuel Availability
EXP	Express Lane Access
POW	Power

CHAPTER 1

INTRODUCTION

Bayes estimators generally do not require maximization and hence are attractive for problems involving non-convex optimization and for weakly identified models. Discrete choice models widely used in travel behavior modeling include the nested logit and multinomial probit models, which involve maximizing a non-concave likelihood function (McFadden 2001; Brownstone 2001; Train 2009). In addition, frequentist inference on willingness-to-pay (WTP) for qualitative improvements – one of the most important outputs from discrete choice analysis, including measures such as the value of time– is limited due to weak identification (see Bolduc et al. 2010). Finally, the asymptotic properties of the Bayes estimator are better than those of the maximum simulated likelihood estimator that is necessary for complex models such as the mixed logit model, basically because the Bayes estimator avoids the bias introduced by simulation of the loglikelihood function.

Despite the advantages of Bayes estimators, the dominant approach in travel behavior modeling is frequentist estimation. One of the common arguments against Bayesian econometrics is the effect of priors. Not only the effect of prior distributions disappears as the prior precision gets lower, but also the effect of the prior vanishes when sample sizes are large enough (under regularity conditions that apply in preference space). Furthermore, for small samples and for models with weakly identified parameters, prior distributions may help to obtain better results. A second argument against simulation-based Bayes estimators is that there are no general rules for assessing convergence (cf. Cowles and Carlin 1996) to the posterior distribution of interest. Additionally, there is concern about correlation in the simulated chains that are used to generate Bayes estimators.

In this thesis, through two empirical analysis of travel mode choice, we study convergence of the Markov chain Monte Carlo (MCMC) estimator of several discrete choice models (logit and probit, in preference and willingness-to-pay spaces; cf. Godoy and Ortzar 2008). The contribution of this thesis is to establish general, applied guidelines on the number of iterations that are required for an average model of mode choice, as well as on the most suitable diagnostic test for convergence for choice problems in transportation analysis. Although there are examples of convergence analysis in marketing, applications in transportation have special characteristics. Sample sizes are usually larger, data is disaggregated, and the role of alternative-specific attributes (rather than just individual-specific characteristics) is fundamental. Different convergence tests are illustrated and compared in this work in an effort to review the different tools that are available for assessing convergence of Bayes estimators of not only taste parameters (preference space) but also willingness-to-pay. Motivated by the fact that the asymptotic properties of the Bayes and frequentist estimators coincide (under regularity conditions), another contribution is our comparison of the posterior distributions of interest with the results of the maximum likelihood estimation.

The rest of the thesis is organized as follows. In chapter 2, we review the Gibbs sampler of the multinomial probit model, Metropolis-Hastings of the multinomial logit model as well as the different diagnostics for convergence that have been proposed in the literature. We focus in the most popular tests, including the Geweke test of nonstationarity, the Raftery-Lewis diagnostic, and the Gelman-Rubin test. In chapter 3 and chapter 4 we briefly describe the two independent survey data that we use for the empirical studies of convergence. Meanwhile we apply these tests to the chains generated for estimation of a multinomial probit model with full covariance matrix of travel mode choice. Finally, chapter 5 concludes and provides insights for further research.

CHAPTER 2

MCMC ESTIMATORS AND CONVERGENCE ANALYSIS

The goal of Bayesian inference is to examine the posterior distribution $p(\theta \mid y, x)$. In the case of discrete choice models, θ are the parameters of the random utility $\mathrm{U}_n = X_n\beta_n + \varepsilon_n$, where β_n are marginal utilities that may be individual-specific and X_n is a design matrix that contains the attribute levels as experienced by individual *n*. Note that θ may contain nuisance parameters of the error term ε_n. y is a vector of choice indicators.

2.1 Gibbs sampling and Metropolis-Hastings for discrete choice models

In general, Bayes estimators of discrete choice models are based on simulation-aided inference through the use of Markov chain Monte Carlo (MCMC) methods. This is due to the fact that the general posterior distribution is not known explicitly. MCMC methods build a stochastic sampler using a Markov chain with the posterior of interest as its equilibrium distribution. The most common MCMC methods are the Gibbs and the Metropolis-Hastings samplers.

For implementation of the Gibbs sampler, consider the partition B of $\theta \in \Theta$ such that $\theta' = \left(\theta'_{(1)}, \theta'_{(2)}, \ldots, \theta'_{(\mathrm{B})}\right)$. For every subvector $\theta'_{(\mathrm{p})}$, consider $\theta'_{<(p)} = \left(\theta'_{(1)}, \theta'_{(2)}, \ldots, \theta'_{(p-1)}\right)$, $\theta'_{>(p)} = \left(\theta'_{(p+1)}, \theta'_{(p+2)}, \ldots, \theta'_{(\mathrm{B})}\right)$, $\theta'_{<(1)} = \{\varnothing\}$, $\theta'_{>(\mathrm{B})} = \{\varnothing\}$ and $\theta'_{-(p)} = \left(\theta'_{<(p)}, \theta'_{>(p)}\right)$. Assume that the full conditional distributions $\pi\left(\theta'_{(\mathrm{p})} \mid \theta'_{-(p)}\right)$ have all a known closed-form. Then, the transition process for $p(\theta \mid y, x)$ from $\theta^{(g-1)}$, in iteration *(g-1),* to $\theta^{(g)}$, in iteration *(g),* is given by $\theta_{(p)}^{(g)} \sim \pi\left(\theta'_{(p)} \mid \theta'^{(g)}_{<(p)}, \theta'^{(g-1)}_{>(p)}, y, X\right) \quad, \forall p \in \mathrm{B}$. Iterative sampling creates a reversible Markov chain with elements that are drawn from the desired posterior distribution $\theta_{(p)}^{(g)} \sim p(\theta \mid y, x) \quad, \forall g$.

Albert and Chib (1993) proposed a Gibbs sampler for the binary probit model, which was extended by McCulloch et al. (2000) to the multinomial probit. The sampler exploits the fact of the utility function being normally distributed with a truncation region determined by the choice indicators. Consider the multinomial probit model $\mathrm{U}_n = X_n\beta_n + \varepsilon_n$, where the error term is a multivariate normally distributed vector of dimension *J* (total number of alternatives) with full covariance matrix Σ , i.e. $\varepsilon_n \sim \mathrm{N}(0, \Sigma_{J\times J})$. After setting location using utility differences with respect to alternative *j*, the estimable form of the model is $\Delta_j \mathrm{U}_n = \Delta_j X_n \beta_n + \Delta_j \varepsilon_n$, $\Delta_j \varepsilon_n \sim N\left(0_{(J-1)}, \Delta_j \Sigma \Delta'_j\right)$. (Δ_j is a matrix difference operator.) The multinomial probit Gibbs sampler uses the partition $\{\beta, \Sigma, \Delta_j U\}$. Note that the parameter space is augmented in the unobservable random utility. Iterations of the sampler are summarized as follows:

(1) Start with $\beta^{(g-1)}$, $\Delta_j U^{(g-1)}$ and $\Sigma^{(g-1)}$. (The starting $\beta^{(0)}$, $\Delta_j U^{(0)}$ and $\Sigma^{(0)}$ can be set at any value.)

(2) Condition on $\beta^{(g-1)}$ and $\Sigma^{(g-1)}$ and for every individual n sample a new value for the latent random utility (in differences) from the distribution

$$\Delta_j U^{(g)} \sim \begin{cases} \mathrm{N}\left(\Delta_j X_n \beta, \Delta_j \Sigma \Delta_j'\right) \mathbb{1}\left(\Delta_j U_{in} < 0, \forall j \neq i\right) \\ \mathrm{N}\left(\Delta_j X_n \beta, \Delta_j \Sigma \Delta_j'\right) \mathbb{1}\left(\Delta_j U_{in} > \max\left\{0, \Delta_j U_{i,-j}\right\}, \forall j \neq i\right) \end{cases} \tag{1}$$

The latent utilities are N realizations of a truncated normal distribution with known mean and variance, given step 1.

(3) Using the samples from step 2, condition on the vector $\Delta_j U^{(g)}$ built by stacking $\Delta_j U_n^{(g)}$, $\forall n$. Updated elements of the marginal utilities are drawn from

$$\begin{aligned} \beta^{(g)} \sim N\Bigg(& \left(\breve{V}_\beta^{-1} \breve{\beta} + \left(C^{(g-1)'} X\right)' C^{(g-1)'} X\right)^{-1} \left(\breve{V}_\beta^{-1} + X' C^{(g-1)} C^{(g-1)'} \Delta_j U^{(g)}\right) \quad , \\ & \left(\breve{V}_\beta^{-1} + C^{(g-1)'} X' \left(C^{(g-1)'} X\right)\right)^{-1} \Bigg) \end{aligned} \tag{2}$$

where $\breve{\beta}$ and $\breve{V}_\beta^{-1}$ are the parameters of a Gaussian prior distribution $p(\beta) \sim N\left(\breve{\beta}, \breve{V}_\beta\right)$, and where $C^{(g-1)}$ is the Cholesky root of the covariance matrix $\Delta_j \Sigma^{(g-1)} \Delta_j'$.

(4) Update the covariance matrix

$$\left(\Delta_j \Sigma \Delta_j'\right)^{(g)} \sim IW\left(\breve{\upsilon} + N, \Delta_j \breve{\Sigma} \Delta_j' + \sum_{i=1}^{N} \Delta_j \varepsilon_i \varepsilon_i' \Delta_j'\right) \quad \middle| c_{11} = 1 \tag{3}$$

where the $\breve{\upsilon}$ and $\Delta_j \breve{\Sigma} \Delta_j'$ are the parameters of the inverted-Whishart prior $p\left(\Delta_j \Sigma \Delta_j'\right) = IW\left(\breve{\upsilon}, \Delta_j \breve{\Sigma} \Delta_j'\right)$.

(5) Update $g=g+1$, and go back to step 1.

The Gibbs sampler sequence of iterative random draws forms asymptotically an irreducible, recurrent, aperiodic and therefore ergodic Markov chain that converges at an exponential rate to the desired posterior distribution. Although the multinomial probit Gibbs sampler will always exhibit asymptotic convergence to the posterior distribution of interest, a nontrivial empirical problem is the definition of a sufficiently large number of repetitions for practical implementation of the sampler. In fact, the solution to this problem becomes less straightforward as the posterior distribution of the parameters of complex models may be multimodal.

Metropolis-Hastings (M-H) is needed when there is at least one element in the parti tion without a closed-form conditional distribution. In this case, the transition process cannot exploit the full conditional distributions and a transition process is needed. For instance, a new candidate can be generated from a (Gaussian) random walk. Rossi et al. (2005) introduce an independent Metropolis-Hastings for the multinomial logit model that works as follows:

(1) Start with $\beta^{(g-1)}$, the starting value $\beta^{(0)}$ are arbitrary.

(2) Condition on y and X, we maximize log-likelihood function which has a closed form for mutinimial logit model to get estimates $\hat{\beta}_{MLE}$.

(3) Use $N\left(\breve{\theta}, \breve{T}^{-1}\right)$ as prior of θ , proposal distribution as $\beta^{cand} \sim MSt\left(\upsilon, \hat{\beta}_{MLE}, S\left[\breve{T} + I\left(\hat{\beta}_{MLE}\right)\right]^{-1} S\right)$. where S and υ are tuning parameters that can be adjusted to obtain an ideal acceptance rate (see discussion below).

(4) Calculate the acceptance rate.

$$\alpha = \min\left\{1, \frac{l\left(\theta^{cand}; y, I \mid X\right) p\left(\theta^{cand}\right) q\left(\theta^{cand} \mid \theta^{curr}\right)}{l\left(\theta^{curr}; y, I \mid X\right) p\left(\theta^{curr}\right) q\left(\theta^{curr} \mid \theta^{cand}\right)}\right\} \tag{4}$$

(5) According to the acceptance rate, update $\beta^{(g)}$.

(6) Update $g = g + 1$, and go back to step 1.

The M-H sampler above is based on the following asymptotic approximation to the posterior distribution of the multinomial logit model (Scott 2003):

$$p(\beta \mid y, X) \propto \left|I(\beta)\right|^{\frac{1}{2}} \exp\left(\frac{1}{2}\left(\beta - \hat{\beta}_{MLE}\right)' I(\beta)\left(\beta - \hat{\beta}_{MLE}\right)\right) \tag{5}$$

where $\hat{\beta}_{MLE}$ is the maximum likelihood estimator of β, *i.e*. the value $\hat{\beta}_{MLE}(y \mid X)$ that maximizes the likelihood function $l(\beta; y \mid X)$ once y is observed, and where $I(\beta)$ is the Fisher information matrix of the multinomial logit model. (Instead of considering the maximum likelihood estimator, the approximation equation (4) can also be evaluated at the posterior mode (Chib et al. 1998).)

The proposal distribution in step (3) of the M-H sampler above is an example of *independence* Metropolis. Based on the asymptotic approximation to the posterior in equation (5), Rossi et al. (2005) actually proposed two transition processes for updating β. For a *random-walk* Metropolis chain, the candidate realization is defined as $\beta^{cand} = \beta^{curr} + \varepsilon$, where $\varepsilon \sim N\left(0, s^2 I^{-1}\right)$ and s^2 is the precision. For an *independence* Metropolis, the candidate realization is found using $\beta^{cand} \sim Mst\left(\upsilon, \hat{\beta}_{MLE}, I^{-1}\right)$, *i.e.* β^{cand} is drawn from a multivariate t distribution with mean $\hat{\beta}_{MLE}$, dispersion I^{-1}, and υ degrees of freedom. (This is a generalization of the estimator proposed by Chib et al. (1998).)

Consider $p(\beta) \sim N\left(\breve{\beta}, \breve{T}^{-1}\right)$, a multivariate normal prior on β with mean $\breve{\beta}$ and precision $\breve{T}$. To include the effect of the prior precision it is possible to extend both transition processes. Thus, it is possible to consider the following general transition process $\beta^{cand} \sim N\left(\beta^{curr}, S\left[\breve{T} + I\left(\hat{\beta}_{MLE}\right)\right]^{-1} S\right)$ for the random walk, and $\beta^{cand} \sim Mst\left(\upsilon, \hat{\beta}_{MLE}, S\left[\breve{T} + I\left(\hat{\beta}_{MLE}\right)\right]^{-1} S\right)$ for independence Metropolis-Hastings, where S is a diagonal matrix with elements that adjust the covariance matrix of the candidate to get satisfactory acceptance ratios.

We will discuss now the implementation of the *slice sampler* (Wakefield et al. 1991; Damien et al. 1999; Neal 2003), which is a special case of the Metropolis-Hasting algorithm. The slice sampler exploits data augmentation methods, where auxiliary variables are summed up to the parameter space in order to get full conditional distributions of standard form while keeping the marginal posterior of interest. Consider the posterior distribution of the parameters of the multinomial logit model, which depends on the choice probabilities through the likelihood function as well as on the prior. When introducing the auxiliary variables of the slice sampler, the likelihood function can be rewritten as an associated completion where the choice probabilities truncate the parameter space to reflect information summarized by the choice indicators. Then, with a normally distributed prior, the posterior can be simulated from a truncated multivariate normal distribution.

In effect, consider a Bayesian multinomial logit model with prior $p(\beta) \sim N(\breve{\beta}, \breve{T}^{-1})$. Note that the choice probabilities $P_{in} = P\left(\varepsilon_{jn} - \varepsilon_{in} < \left(X_{in} - X_{jn}\right)' \beta \right)$ $, \forall j \neq i \in C_n$ are given by the cumulative distribution of a multivariate logistic. In the specific case of binary choice, the choice probability of alternative i is $P_{in} = P\left(\varepsilon_{jn} - \varepsilon_{in} < \left(X_{in} - X_{jn}\right)' \beta \right)$. Even though it is possible to write a similar expression for P_{jn}, note that in the binary case $P_{jn} = 1 - P_{in}$. Since the difference $\varepsilon_{jn} - \varepsilon_{in}$ for a normalized model has a standard logistic distribution, then $P_{in} = E(y_{in} \mid X_{in}) = \Lambda\left[\left(X_{in} - X_{jn}\right)' \beta \right]$, where $\Lambda\left[\left(X_{in} - X_{jn}\right)' \beta \right] = \exp\left(\left(X_{in} - X_{jn}\right)' \beta \right)$. $/\left(1 + \exp\left(\left(X_{in} - X_{jn}\right)' \beta \right)\right)$ .Because $P_{jn} = 1/\left(1 + \exp\left(\left(X_{in} - X_{jn}\right)' \beta \right)\right)$, then the posterior takes the form

$$p(\beta \mid y, X) \propto p(\beta) \prod_{n=1}^{N} \left[\frac{\exp\left(\left(X_{in} - X_{jn}\right)' \beta \right)}{1 + \exp\left(\left(X_{in} - X_{jn}\right)' \beta \right)} \right]^{y_{in}} \left[\frac{1}{1 + \exp\left(\left(X_{in} - X_{jn}\right)' \beta \right)} \right]^{y_{in} - 1} \tag{6}$$

Consider now the following uniform auxiliary variable

$$u_n = U\left(0, \frac{\exp\left(\left(X_{in} - X_{jn}\right)' \beta \right)}{1 + \exp\left(\left(X_{in} - X_{jn}\right)' \beta \right)} \right), \tag{7}$$

where the logit expression has been conveniently rewritten to summarize both P_{in} and P_{jn}.

Using the auxiliary variable u_n, the posterior (6) becomes

$$p(\beta \mid y, X) \propto p(\beta) \prod_{n=1}^{N} 1_{u_n \leq \frac{\exp\left(\left(X_{in} - X_{jn}\right)' \beta\right)}{1+\exp\left(\left(X_{in} - X_{jn}\right)' \beta\right)}} , \tag{8}$$

Thus, the slice sampling estimator is built from iterative draws inside a Gibbs sampler, which at the g^{th} iteration considers

$$u_n^{(g)} = U\left(0, \frac{\exp\left(\left(X_{in} - X_{jn}\right)' \beta^{(g)}\right)}{1+\exp\left(\left(X_{in} - X_{jn}\right)' \beta^{(g)}\right)}\right), \forall n$$

$$\beta^{(g+1)} \sim N\left(\breve{b}, \breve{T}^{-1}\right) \prod_{n=1}^{N} 1_{u_n^{(g)} \leq \frac{\exp\left(\left(X_{in} - X_{jn}\right)' \beta^{(g)}\right)}{1+\exp\left(\left(X_{in} - X_{jn}\right)' \beta^{(g)}\right)}} ,$$

i.e. β is simulated from a multidimensional truncated normal distribution.

If $\breve{T}$ is assumed to be diagonal, then element β_k is simulated from

$$\beta_k^{(g+1)} \sim N\left(\breve{\beta}_k, \breve{\sigma}_k^2\right) \prod_{n=1}^{N} 1_{u_n^{(g)} \leq \frac{\exp\left(\left(X_{in} - X_{jn}\right)' \beta^{(g)}\right)}{1+\exp\left(\left(X_{in} - X_{jn}\right)' \beta^{(g)}\right)}} , \tag{9}$$

where $\breve{\beta}_k$ and $\breve{\sigma}_k^2$ are hyperparameters of the normally distributed prior. Note that the conditional distribution of θ_k depends on the vectors $\beta_{<k}^{(g+1)'} = \left(\beta_1^{(g+1)'}, \ldots, \beta_{k-1}^{(g+1)'}\right)$ and $\beta_{>k}^{(g+1)'} = \left(\beta_{k+1}^{(g+1)'}, \ldots, \beta_K^{(g+1)'}\right)$, which are both implicitly summarized by $\beta^{(g)}$.

The truncation region is defined by the choice indicator. If $y_{in} = 1$, then $u_n \leq \Lambda\left[\left(X_{in} - X_{jn}\right)' \beta\right]$ which can be inverted into $\left(X_{in} - X_{jn}\right)' \beta \geq \ln\left(u_n / (1 - u_n)\right)$. If $y_{in} = 0$, then $\left(X_{in} - X_{jn}\right)' \beta \leq -\ln\left(u_n / (1 - u_n)\right)$. Then, the truncation region can be summarized as

$$\bigcap_{n=1}^{N}\left\{\beta:(-1)^{y_{in}}\left(X_{in}-X_{jn}\right)'\beta\geq\ln\left(\frac{u_n}{1-u_n}\right)\right\} \tag{10}$$

For each parameter, the truncation region (10) yields the following thresholds

$$\underline{\beta}_k=\max_{n:y_{in}=1}\left\{\left(x_{kin}-x_{kjn}\right)^{-1}\left(\ln\frac{u_n}{1-u_n}-\left(x_{in}-x_{jn}\right)'\beta_{\backslash k}\right)\right\}$$

$$\bar{\theta}_k=\min_{n:y_{in}=1}\left\{\left(x_{kin}-x_{kjn}\right)^{-1}\left(\ln\frac{1-u_n}{u_n}-\left(x_{in}-x_{jn}\right)'\theta_{\backslash k}\right)\right\}$$

where $\underline{\beta}_k\leq\beta_k\leq\bar{\beta}_k$ and $\beta_{\backslash k}=\left(\beta_1,...\beta_{k-1},0,\beta_{k+1},...,\beta_K\right)'$. Then, in the slice sampler it is possible to consider $\beta_k\sim N_{\underline{\beta}_k\leq\theta_k\leq\bar{\beta}_k}\left(\breve{\beta}_k,\breve{\sigma}_k^2\right)$. This is the procedure that Winbugs uses for sampling the parameters of a binary logit. The method can be easily extended to accommodate multinomial choice.

Using the Metropolis-Hastings above as kernel, it is possible to derive a Metropolis-Hastings estimator of a multinomial logit model with random parameters (mixed logit).

2.2 Gibbs sampling and Metropolis-Hastings for discrete choice models

Both the Gibbs sampler and Metropolis-Hastings sequences of iterative random draws asymptotically build an irreducible, recurrent, aperiodic and therefore ergodic Markov chain that converges at an exponential rate to the desired posterior distribution (the Bayesian counterpart of the asymptotic distribution of the parameters of interest). Although both the multinomial probit Gibbs sampler and the M-H logit sampler will always exhibit asymptotic convergence to the posterior distribution of interest, a nontrivial empirical problem is the definition of a sufficiently large number of repetitions for practical implementation of the sampler. In fact, the solution to this problem becomes less straightforward as the posterior distribution of the parameters of complex models may be multimodal. Furthermore, the asymptotic distributions of the parameters in preference and willingness-to-pay spaces do not behave equally. Due to the potential problem of weak identifications that is encountered for willingness-to-pay, the asymptotic posterior distribution is not necessarily normal.

A first tool for assessing convergence is **visual inspection** of the simulated chains. Whereas visual inspection allows the researcher to informally assess how well the chain is covering the parameter space, the literature proposes a variety of formal diagnostic tests. However, it is not possible to single out the best diagnostics. As pointed out by Lahiri and Gao (2002), tests based on a single chain -- either a short pilot chain or a very long chain -- and on multiple short chains are currently the dominant methods. In the following paragraphs we summarize the most common diagnostics, emphasizing the intuition behind the mathematical derivation of each test. Technical details are given in the original papers proposing each test.

The **Geweke test of nonstationarity** (Geweke 1992) looks at the means of two different portions of the sampled values of a single MCMC output. The difference of the means divided by the estimated standard error is used as statistic. Under the null hypothesis of no difference in the separate means (stationarity), the statistic is asymptotically standard normally distributed. Geweke recommends that the initial portion be built using the first 10% of the chain (after the burn-in period), and that the final subsequence be built using the last half of the chain. Although the Geweke test gives insights about stationarity of the finite chain, it focuses only on convergence of the mean and it does not answer questions about the necessary number of iterations or the length of the burn-in period.

The **Raftery-Lewis diagnostic** test (Raftery and Lewis 1992) looks at the behavior of the posterior distribution with respect to a pre-specified quantile to determine whether the MCMC output performs as a Markov chain or not. The Raftery-Lewis test calculates different estimates, including the number of iterations (*total length*) and the burn-in required to ensure that each single series behaves as a Markov chain for the target quantile, level of accuracy, and minimum probability. Another estimate is the minimum number of iterations (*lower bound*) for achieving the given accuracy when estimating the quantile of interest under the assumption of zero autocorrelation. The presence of positive posterior autocorrelation increases the required run length, as summarized by a *dependence factor* that calculates the necessary relative increase. Whereas a dependence factor that equals one indicates independence, values above five indicate strong effects of starting values, high correlation among parameters, or poor mixing (Cowles and Carlin 1996).

The **Gelman-Rubin test** (Gelman and Rubin 1992) is based on asymptotic normality of the posterior distribution and is constructed using a series of independent sequences with over-dispersed starting values. A convergence diagnostic is built using within-chain and between-chain to determine the potential scale reduction factor by which scale might shrink for infinite sampling. Since parallel runs are needed, the Gelman-Rubin test is not computationally efficient. Another problem is the assumption of normality. Whereas normality is a valid assumption for large samples as the Bayes estimators are asymptotically normal, normality is not ensured for small samples.

There are good reviews and assessment of these tests in the literature, including the work of Cowles and Carlin (1996) and Brooks and Roberts (1997). However, most of the empirical work is applied to marketing data and in the few examples of Bayes estimators applied to transportation demand in general there is no convergence analysis. Microdata on travel demand has special characteristics: sample sizes are usually larger (with a decreasing importance of prior distributions), data is disaggregated, and alternative-specific attributes (rather than just individual-specific characteristics) are central. An exception to the lack of testing convergence in travel behavior analysis is the work of Godoy and Ortzar (2008). The authors proposed an interesting iterative procedure based on the Geweke, Raftery-Lewis, and **Heidelberger-Welch** tests. The problem with the proposed procedure is that it is focused on thinning the MCMC chain. Thinning an MCMC chain means to discard consecutive samples of the chain with the goal of reducing autocorrelation. A thinned Markov chain keeps every k^{th} iteration, where k is known as thinning parameter. Thinning a Markov chain was standard practice years ago, basically to save computer space for storage of the chain as well as to save computing time for post-processing (plotting and make calculations) very long runs. Saving computer space is no longer an issue. Thinning may actually be computationally inefficient, because the total length of the run is amplified by the thinning factor k, then a $(k\text{-}1)/k$ fraction of the chain is considered useless and erased. Most likely equally precise results could have been obtained by considered a shorter chain without thinning. In fact, MacEachern and

Berliner (1994) and Link and Eaton (2012) show that posterior inference is more precise when the entire Markov chain is used. Even more, autocorrelation is not a real problem in Bayesian inference. Whereas the general rule is that Monte Carlo integration works when the simulated samples are iid, Monte Carlo integration also works with dependent samples if these correlated samples form a chain that is ergodic. Under very general conditions, both the Gibbs and Metropolis-Hastings samplers form Markov chains that are ergodic.

CHAPTER 3

WILLINGNESS-TO-PAY: POST-PROCESSING AND DIRECT INFERENCE

In a discrete choice model, the parameter of economic (and policy) interest is a vector of willingness-to-pay, not the marginal utilities (which have an unknown measurement scale). Willingness-to-pay (WTP) is a marginal rate of substitution that measures the maximum amount of money individuals would pay for a unit improvement of a particular attribute. Because the parameter space of standard discrete choice models directly considers marginal utilities (preference space), the derivation of WTP requires making inference on parameter ratios.

Producing robust standard errors, building confidence intervals, and even calculating the mean of parameter ratios is complicated by the problem of weak identification. Weak identification happens when the likelihood function has relatively flat areas in the parameter space of interest. With weakly identified WTP parameters, statistical properties of the WTP estimators (such as consistency and convergence) may not be guaranteed, even if the estimator of the marginal utilities behaves well. In fact, under weak identification the Bayes and frequentist estimators do not exhibit the same asymptotic properties. As mentioned in the introduction of this thesis, WTP identification is the one of the main motivations for the empirical analysis of convergence that we perform in this thesis.

Consider a discrete choice problem, where individual $n \in \{1,2,\ldots,N\}$ chooses a mutually exclusive alternative $i \in \{1,2,\ldots,J\}$. $U_n = [U_{1n}, U_{2n}, \ldots, U_{Jn}]$ represents the vector of random utilities for the whole choice set. If $X_{in} = (c_{in}, x_{in,1}, \ldots, x_{in,K})'$ is a vector of attributes, where c_{in} is cost of the discrete alternative, assuming a linear specification the random utility can be written as:

$$U_{in} = x_{in}'\beta = c_{in}\beta_C + x_{in,1}\beta_1 + x_{in,2}\beta_2 + \cdots + x_{in,K}\beta_K + \varepsilon_{in}, \tag{11}$$

where β_C is the additive inverse of the marginal utility of income.

In the model above, WTP can be derived by considering the marginal rate of substitution:

$$WTP_k = -\frac{1}{\beta_C}\frac{\partial U_{in}}{\partial x_{in,k}} = -\frac{\beta_k}{\beta_C}, \tag{12}$$

Using the estimates $\hat{\beta}(y \mid X)$, WTP for improvements in attribute k is usually estimated as $-\hat{\beta}_k(y \mid X)/\hat{\beta}_C(y \mid X)$. The asymptotic distribution of this ratio is nonstandard, making it difficult to derive standard errors and confidence intervals. In fact, estimating mean effects can be difficult as the distribution of the ratio may not have finite moments. When the parameters β are random to represent a heterogeneity distribution, the inference problem becomes even more involved.

Due to the issues mentioned above, Train and Weeks (2005) proposed the recast the parameter space for direct inference on WTP. What \cite{Train:05} suggested was to use the nonlinear utility:

$$U_{in} = c_{in}\beta_C + \beta_C\left(x_{in,1}WTP_1 + x_{in,2}WTP_2 + \cdots + x_{in,K}WTP_K\right) + \varepsilon_{in}, \tag{13}$$

where the parameters are now $(\beta_C, WTP_1, \ldots, WTP_K)'$. As a result, WTP and standard errors are calculated directly.

When Bayesian econometrics is used, the output of the estimation is the posterior probability of the structural parameters of the model. If one is using the model in preference space as in equation (11), MCMC estimators produce samples from the posterior $p(\beta \mid y, X)$. These samples can be used to derive the posterior of interest $p(WTP \mid y, X)$ in a technique known as **post-processing**. In effect, the ratio

$$WTP_k^{(r)} = -\frac{\beta_k^{(r)}}{\beta_C^{(r)}} \quad , \qquad (14)$$

where $\beta_k^{(r)}$ and $\beta_C^{(r)}$ are the *r*-th draw in the chain of the simulated posterior $p(\beta \mid y, X)$, is a draw from the desired WTP posterior. Point and interval estimates can be then easily derived from the simulated (post-processed) posterior $p(WTP \mid y, X)$.

The desired posterior $p(WTP \mid y, X)$ can also be derived from direct Bayesian inference, where the MCMC sampler will be constructed from the nonlinear specification in equation (13). In particular, for the probit Gibbs sampler, instead of using $\Delta_j U_n = \Delta_j X_n \beta + \Delta_j \varepsilon_n$ direct Bayesian inference would start with the model rewritten as $\Delta_j U_{in} = \Delta_j c_{in} \beta_C + \Delta_j \beta_C \left(x_{n,1} WTP_1 + x_{n,2} WTP_2 + \cdots + x_{n,K} WTP_K \right) + \Delta_j \varepsilon_{in}$. All steps of the Gibbs sampler need to be updated to simulate draws of the posterior $p(\beta_C, WTP \mid y, X)$.

In the case of the multinomial logit model, the main change in the Metropolis-Hastings sampler is to consider the following likelihood function for the deriving the maximum likelihood estimate (needed as mean of the proposal distribution) and for calculating the acceptance rate:

$$\ell\left(\beta_C, WTP; y \mid X\right) = \prod_{n=1}^{N}\prod_{j=1}^{J} P_{in}^{y_{in}} , \tag{15}$$

where

$$P_{in} = \frac{\exp\left(c_{in}\beta_C + \beta_C\left(x_{in,1}WTP_1 + x_{in,2}WTP_2 + \cdots + x_{in,K}WTP_K\right)\right)}{\sum_{j=1}^{J}\exp\left(c_{jn}\beta_C + \beta_C\left(x_{jn,1}WTP_1 + x_{jn,2}WTP_2 + \cdots + x_{jn,K}WTP_K\right)\right)}. \tag{16}$$

CHAPTER 4

CANADIAN VEHICLE MARKET SURVEY DATA

4.1 Data description

For the empirical analysis of convergence we use the personal vehicle choice data from a survey by EMRG (Energy and Materials Research Group, Simon Fraser University) in 2002 (Horne 2003), which is well used in hybrid choice model using frequentist estimator (Bolduc et al. 2008) and Bayes estimator (Daziano and Bolduc 2013). The survey, under the stated preferences design, did telephone interview for questionnaire on 1500 Canadian consumers in urban area and then mailed to them. There were 866 complete responses (77%) from the survey and 1877 usable data after cleaning (since each respondent provided several choices). The stated preference experiment consisted of 4 choice situations which were standard gasoline vehicle (SGV), alternative fuel vehicle (AFV), hybrid electric vehicle (HEV) and Hydrogen fuel cell vehicle (HFC).

As summarized in Table 4.1, the experimental attributes were: capital cost [CAD2002/10000], monthly operating cost [CAD2002/100-month], fuel availability [ratio], Express lane access and relative power [ratio].

The survey also collected sociodemographic data (age, income, gender, education level, etc), as well as attitudinal indicators (Daziano and Bolduc 2013).

Table 4.1: Summary of descriptive statistics

Attributes	Mean	S.D.	Min	Max
Capital Cost SGV [CAD2002/10000]	2.21	1.06	0.02	8.17
Capital Cost AFV [CAD2002/10000]	2.21	1.05	0.02	7.81
Capital Cost HEV [CAD2002/10000]	2.31	1.11	0.02	8.52
Capital Cost HFC [CAD2002/10000]	2.36	1.13	0.02	8.52
Monthly Operating Cost SGV [CAD2002/100-month]	1.42	0.82	0.11	7.80
Monthly Operating Cost AFV [CAD2002/100-month]	1.42	0.79	0.11	7.20
Monthly Operating Cost SGV [CAD2002/100-month]	1.07	0.62	0.08	5.85
Monthly Operating Cost SGV [CAD2002/100-month]	1.37	0.78	0.12	7.20
Fuel Availability AFV [ratio]	0.50	0.25	0.25	0.75
Fuel Availability HFC [ratio]	0.50	0.25	0.25	0.75
Express Lane Access AFV	0.48	0.50	0.00	1.00
Express Lane Access HEV	0.48	0.50	0.00	1.00
Express Lane Access HFC	0.51	0.50	0.00	1.00
Power Compared to SGV AFV [ratio]	0.95	0.05	0.90	1.00
Power Compared to SGV HEV [ratio]	0.95	0.05	0.90	1.00
Power Compared to SGV HFC [ratio]	0.95	0.05	0.90	1.00

* Note that CAD2002 means Canadian dollar in 2002

4.2 Results

In this section, we test empirical convergence of the MCMC chains of:

(1) the posterior of marginal utilities (preference space)

(2) the post-processed posterior of willingness-to-pay (post-processed WTP)

(3) the posterior of willingness-to-pay using direct inference (WTP space)

In addition, we test independent Metropolis-Hastings, random-walk Metropolis-Hastings and slice sampling for multinomial logit model as well as Gibbs sampling for multinomial probit model.

4.2.1 Point estimates

The first step in an estimation problem is to derive point estimates. However, for a Bayes estimator we need to establish a priori how many iterations will be used for constructing samples of the Markov chains. We ran the Raftery-Lewis diagnostic for a pilot sampler of a predetermined length, given that we wanted to measure the 2.5% quantile of the posterior distribution with an acceptable tolerance of 0.005 and a probability of 0.95 of obtaining an estimate of the desired quantile within the preset accuracy (these values are the standard assumptions for the Raftery-Lewis diagnostic). Results of the diagnostic test are presented in Table 4.2.1 for the probit model in preference space.

Table 4.2.1: Raftery-Lewis diagnostic

	Burn-in	**Total Length**	**Lower Bound**	**Dependence factor**
ASC_{AFV}	10	11,638	3,746	3.11
ASC_{HEV}	15	20,331	3,746	5.43
ASC_{HFC}	8	8,644	3,746	2.31
CC	8	9,251	3,746	2.47
FC	10	23,852	3,746	2.43
FA	35	36,635	3,746	9.78
EXP	5	6,151	3,746	1.64
POW	6	6,443	3,746	1.72

The lower bound corresponds to the length of the pilot sampler, and is interpreted as the total number of iterations of an iid sampler. The burn-in iterations is the number of initial draws that should be erased before summarizing the posterior. The estimates of the burn-in iterations seem low, especially if one looks at the effect of differing starting values in Figure 4.2.1. Low dependence factors come at no surprise given the low autocorrelations detected. The dependence factors amplify the lower bounds to derive an estimate of the total length required. If we take the maximum of the total length, the Raftery-Lewis diagnostic suggests to use about 50,000 iterations, i.e. the longer run we have been analyzing in this case. One of the problems with the diagnostic is that it overestimates the effect of autocorrelation and the estimates for total length are conservative.

Table 4.2.2: Multinomial probit: point estimates in preference space

	MSLE		**3,746**		**50,000**	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
ASC_{AFV}	-3.3523	0.6123	-3.0770	0.5969	-3.3961	0.5867
ASC_{HEV}	-0.8957	0.4312	-0.9478	0.4358	-0.8940	0.4277
ASC_{HFC}	-1.4128	0.5121	-1.4396	0.5080	-1.4024	0.5078
CC: Capital Cost*	-0.3772	0.1328	-0.4216	0.1226	-0.3648	0.1448
FC: Monthly Operating Cost†	-0.3094	0.1311	-0.3513	0.1163	-0.2983	0.1372
FA: Fuel Availability [ratio]	0.7341	0.2139	0.7378	0.1536	0.7336	0.2004
EXP: Express Lane Access	0.0720	0.0381	0.0777	0.0372	0.0701	0.0384
POW: Power [ratio]	1.3690	0.4741	1.5034	0.4324	1.3751	0.4966

*CAD2002/10000
†CAD2002/100-month

Tables 4.2.2 (preference space) and 4.2.3 (WTP) present the point estimates and standard errors derived using the Gibbs sampler of a multinomial probit model with a full covariance matrix and non-informative priors. The probit Gibbs sampler was run using 3,746 (minimum requirement for the Raftery-Lewis diagnostic) and 50,000 iterations in preference space, post-processed WTP, and WTP space. The maximum simulated likelihood estimator (MSLE) of the probit model, in preference space, was found using the GHK recursive transformation of the choice probabilities.

Parameters behave as expected in terms of sign and magnitude. However, WTP estimates behave in a rather unexpected way: statistical significance is deteriorated for post-processing, and for both post-processed WTP and WTP-space estimates the magnitude of the ratio is not constant when increasing the number of iterations.

Table 4.2.3: WTP point estimates and standard errors of a multinomial probit model

	Post-processed WTP				**WTP space**			
	3,746		50,000		3,746		50,000	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
ASC_{AFV}	-9.385	3.865	-12.00	13.08	0.050	2.482	6.522	4.896
ASC_{HEV}	-2.647	1.333	-2.611	1.565	-0.412	1.967	-0.625	2.884
ASC_{HFC}	-4.031	1.743	-5.042	5.717	-6.287	1.948	-10.15	3.159
WTP_{FC}	-0.880	0.363	-0.861	0.432	-0.950	0.397	-1.397	0.600
WTP_{FA}	2.012	0.720	3.121	5.142	1.405	0.581	3.119	1.027
WTP_{EXP}	0.206	0.117	0.212	0.146	0.254	0.157	0.327	0.251
WTP_{POW}	3.996	1.539	4.101	2.007	4.916	1.702	8.440	2.903

Tables 4.2.4 (preference space) and 4.2.5 (WTP) show the point estimates and standard errors of a multinomial logit model. Three different proposal distributions for the Metropolis-Hastings algorithm were used, namely independent MH, random-walk MH, and slice sampling. When constructing pretests for the three proposal distributions we noticed that the Markov chains produced values very close to the maximum likelihood estimates even with a very low number of repetitions. We present results with the minimum number of iterations required by the Raftery-Lewis diagnostic (3,746 iterations). With iterations above 20,000 the estimates of the three MH algorithms were identical to the MLE.

Table 4.2.4: Point estimates and standard errors in preference space of multinomial logit model

	MLE		**Independent MH**		**Random-Walk MH**		**Slice Sampling**	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
ASC_{AFV}	-4.499	0.655	-4.517	0.656	-4.531	0.764	-4.434	0.666
ASC_{HEV}	-1.381	0.632	-1.392	0.636	-1.439	0.647	-1.310	0.642
ASC_{HFC}	-2.104	0.640	-2.115	0.641	-2.158	0.677	-2.031	0.651
CC	-0.857	0.208	-0.854	0.208	-0.812	0.245	-0.860	0.208
FC	-0.827	0.197	-0.838	0.199	-0.826	0.206	-0.830	0.198
FA	1.357	0.185	1.361	0.188	1.344	0.218	1.356	0.185
EXP	0.156	0.069	0.154	0.068	0.154	0.079	0.155	0.069
POW	2.704	0.655	2.715	0.659	2.756	0.717	2.634	0.670

* Note that results for MCMC estimators are under 3,746 iterations since points estimates as well as standard error coincide with MLE with 20,000 iteration for all three MCMC estimators

Table 4.2.5: WTP point estimates and standard errors of multinomial logit model

	Benchmark*		**Independent MH**		**Random-Walk MH**		**Slice Sampling**	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
ASC_{AFV}	-5.662	2.202	-5.694	2.292	-5.917	2.434	-5.547	2.439
ASC_{HEV}	-1.742	1.056	-1.766	1.094	-1.891	1.197	-1.645	1.147
ASC_{HFC}	-2.657	1.307	-2.680	1.355	-2.830	1.440	-2.551	1.437
WTP_{FC}	-1.044	0.458	-1.053	0.469	-1.077	0.839	-1.038	0.477
WTP_{FA}	1.710	0.661	1.714	0.672	1.750	0.889	1.694	0.707
WTP_{EXP}	0.198	0.120	0.194	0.116	0.209	0.140	0.195	0.121
WTP_{POW}	3.397	1.355	3.424	1.479	3.600	1.616	3.294	1.647

* Estimates with 20,000 MCMC iterations

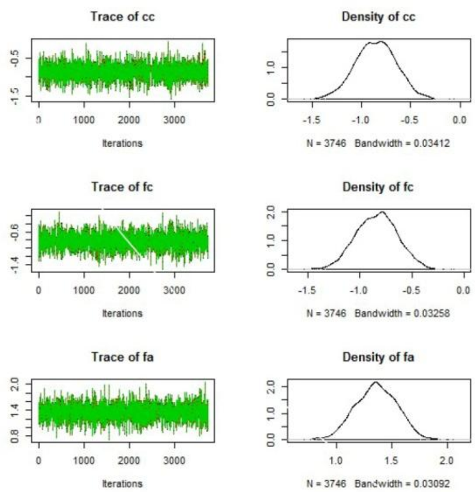
Although the comparison with the MLE provides useful insights about the properties of the posterior distribution, the analysis of convergence should be based on the Bayesian results alone. Each MCMC estimator was rerun three different times -- with different starting values -- to perform the following series of standard convergence tests.

4.2.2 Visual inspection

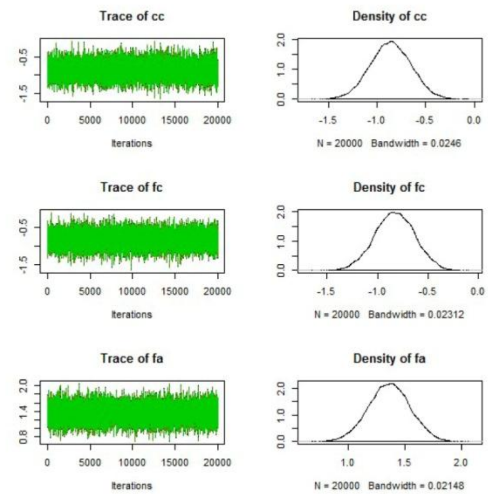
The first test for assessing convergence is visual inspection of traceplots, which display the sampled draws at every iteration for every parameter. For preference space, all parameters behave as white noise for all estimators, indicating quick convergence. However, WTP estimates in general failed to pass the visual inspection test. For instance, Figure 4.2.1 (Appendix) summarizes the resulting traceplots for selected parameters of the multinomial logit model in preference and WTP space, for 3,746 and 20,000 iterations of the independent Metropolis-Hastings estimator. The traceplots of marginal utility show good behavior after 3,746 iterations, when the mean stabilizes and there are no indications of dependence on starting values. Any irregularity in the traceplots (slope and curvature, unstable means, etc.) reflects convergence failure. For example, the traceplots of WTP estimates indicate poor convergence since there exist outliers and the entire chain fluctuates around the posterior mean in an unstable fashion.

It is generally argued that traceplots mask the potential correlation among the different draws that are generated for estimation. Thus, one needs to check the autocorrelation of posterior chains, which usually complements the analysis of traceplots in applied work. In this case, autocorrelations fall quickly to zero before 10 lags, which strengthen accuracy of visual inspection.

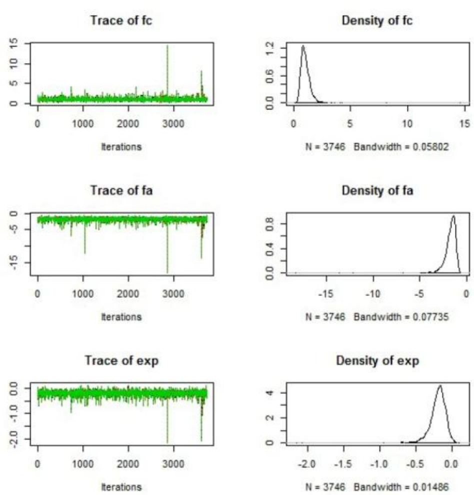
Visual inspection can only provide intuitive information for assessing convergence. In the subsections below we discuss the results of formal convergence tests.



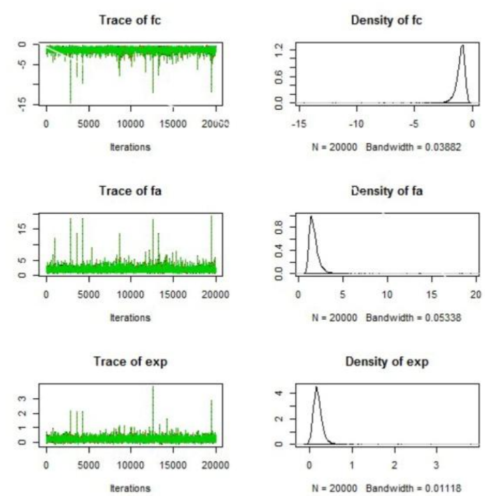
(a) 3,746 iterations for marginal utility



(b) 20,000 iterations for marginal utility



(c) 3,746 iterations for WTP



(d) 20,000 iterations for WTP

Figure 4.2.1: Traceplots and nonparametric density estimates of selected parameters

4.2.3 Geweke test of nonstationarity

The Geweke test compares the mean of two independent portions of the chain by calculating a Z-score statistic. The null hypothesis is equality of the means, and the test statistic is simply calculated as the difference of the sample means divided by the standard error (adjusted by autocorrelation). The asymptotic distribution of the test is a standard normal distribution. Table 4.2.6 and Table 4.2.7 present the Z-scores for the first of three independent chains of each estimator under consideration.

Table 4.2.6: Z-scores of marginal utility for the Geweke test

	Probit model		**Logit model**					
	Gibbs Sampling		Independent MH		Random-Walk MH		Slice Sampling	
	3,746	50,000	3,746	20,000	3,746	20,000	3,746	20,000
ASC_{AFV}	2.813*	-0.534	0.724	-0.384	1.170	-0.883	4.566*	1.370
ASC_{HEV}	-3.801*	-0.352	0.849	-0.862	1.629	-0.757	4.332*	1.545
ASC_{HFC}	-4.010*	-0.254	0.756	-0.554	1.544	-0.772	4.350*	1.476
CC	-3.627*	1.800	0.356	0.247	0.476	1.246	-1.451	-0.206
FC	-2.367*	0.177	1.099	-0.971	0.454	-0.290	1.613	1.381
FA	1.739	1.360	1.085	-0.774	-0.649	0.612	1.036	0.903
EXP	-0.431	1.058	0.413	-1.776	-0.758	-0.101	2.630*	1.404
POW	5.083*	0.173	-1.046	0.774	-1.522	0.678	4.500*	-1.415

* Reject equality of the means at the 95% confidence level.

For the short runs (3,746 iteration) the Z-scores are relatively high. In fact, some of the Z-scores are larger than the critical value at the 95% confidence level, providing no evidence in favor of the null hypothesis (i.e. equality of the posterior means, which means that convergence has been achieved). For the long runs in preference space (50,000 iterations for probit and 20,000 for logit) all estimators pass the Geweke test for all parameters. This is still true for the multinomial logit model in WTP space. However, the probit model in WTP space fails to pass the Geweke test independently of the number of iterations being used. We note that the probit post-processed WTP estimates passed the visual inspection test, which clearly contradicts the results of the Geweke test.

Table 4.2.7: Z-scores of WTP for the Geweke test

	Probit model				**Logit model**					
	Gibbs Sampling		WTP space		Independent MH		Random-Walk MH		Slice Sampling	
	3,746	50,000	3,746	50,000	3,746	20,000	3,746	20,000	3,746	20,000
ASC_{AFV}	-3.629*	-5.174*	23.79*	17.98*	-0.939	-0.632	-0.837	1.086	-2.893*	-1.226
ASC_{HEV}	-1.955	-3.560*	10.60*	12.44*	-1.434	0.570	-1.483	1.025	-1.531	-1.040
ASC_{HFC}	-2.177*	-3.110*	-12.20*	-14.79*	-1.239	0.273	-1.089	1.096	-2.303*	-1.034
WTP_{FC}	1.167	1.575	-8.153*	-8.061*	-1.483	-0.573	-0.807	0.616	-1.204	-1.632
WTP_{FA}	2.275*	2.856*	27.87*	8.104*	0.578	-0.421	1.273	-0.980	1.209	0.605
WTP_{EXP}	0.968	1.219	12.09*	7.752*	0.763	0.661	-1.149	-0.164	0.501	1.303
WTP_{POW}	1.633	1.820	17.72*	15.42*	1.254	-0.319	1.821	-1.118	2.472*	1.048

* Reject equality of the means at the 95% confidence level.

4.2.4 Gelman-Rubin diagnostic

The idea behind the Gelman-Rubin diagnostic is very intuitive. Different chains are built using over-dispersed starting values. Convergence is achieved when the outputs of the chains cannot be distinguished. An indicator of undistinguishable chains is given by potential scale reduction factors. Empirically, the 95% credible interval of the potential scale reduction factors is calculated to account for the uncertainty introduced by chains with finite lengths. Since lack of convergence will produce wide credible intervals, rather than just looking at the point estimates, the diagnostic looks at the 97.5% quantile (the upper bound of the 95% central credible interval).

Table 4.2.8: Gelman-Rubin potential scale reduction factors

	WTP				**WTP space**			
	3,746		50,000		3,746		50,000	
	Est.	Upper C.I.	Est.	Upper C.I.	Est.	Upper C.I.	Est.	Upper C.I.
ASC_{AFV}	1.45	4.18	1.64	2.57	1.76	2.90	1.41	2.31
ASC_{HEV}	1.13	1.38	1.08	1.16	2.05	3.48	1.03	1.21
ASC_{HFC}	1.47	4.41	1.30	1.83	1.01	1.01	1.29	1.78
WTP_{FC}	1.05	1.17	1.04	1.12	1.17	1.49	1.11	1.32
WTP_{FA}	1.51	5.28	1.29	1.79	1.07	1.23	1.09	1.25
WTP_{EXP}	1.08	1.10	1.01	1.01	1.14	1.42	1.08	1.30
WTP_{POW}	1.11	1.24	1.06	1.18	1.04	1.13	1.20	1.57

In practice it is common to use a threshold of 1.2 for the upper bound of the credible interval. Both point estimates and upper bounds of the credible interval of the Gelman-Rubin potential scale reduction factors for most of the estimates in preference space are very close to 1, indicating excellent convergence. Excellent convergence was diagnosed also for the multinomial logit model in WTP-space for the longer runs of the three estimators. However, the results of the Gelman-Rubin diagnostic for the multinomial probit model in WTP-space (Table 4.2.8) indicate converge failure. In sum, the results given by Gelman-Rubin diagnostic coincide with those found with the Geweke test.

CHAPTER 5

SWISSMETRO SURVEY DATA

5.1 Data description

For the empirical analysis of convergence we use the microdata on interurban travel mode choice (Bierlaire et al. 2001), a dataset collected in Switzerland in 1998 that has been well documented and analyzed, using frequestist estimators of discrete choice, e.g. Hess et al. (2005), Bierlaire et al. (2008) and Fosgerau and Bierlaire (2009). (All the features of the Biogeme software for frequentist analysis of discrete choice models are exemplified using the Swissmetro data.) The data, based on stated preferences, comprises two samples: train users (441 valid respondents) and car users (750 valid respondents). The stated preference experiment consisted of nine hypothetical choice situations where the experimental alternatives were car (for car owners only), train, and Swissmetro. The goal of the survey was to help determine the potential demand for an innovative underground, high-speed transportation system. Swissmetro was a project being evaluated at that time by the Swiss government that was never built and went into liquidation in 2009 due to lack of support.

As summarized in Table 5.1, the experimental attributes were travel time [min], travel cost [CHF], headway for train and Swissmetro [min], and the quality of the seats for Swissmetro (airline-like seat configuration or not). The survey also collected

Table 5.1: Summary of descriptive statistics for the Swissmetro data

Attributes	**Mean**	**S.D.**	**Min**	**Max**
Travel Time Train [min]	166.63	77.35	31.00	1049.00
Travel Time Swissmetro [min]	87.47	53.55	8.00	796.00
Travel Time Car [min]	123.80	88.71	0.00	1560.00
Travel Cost Train [CHF]	514.34	1088.93	4.00	5040.00
Travel Cost Swissmetro [CHF]	670.34	1441.59	6.00	6720.00
Travel Cost Car [CHF]	78.74	55.26	0.00	520.00
Headway Train [min]	70.10	37.43	8.00	796.00
Headway Swissmetro [min]	20.02	8.16	10.00	30.00

sociodemographic data (age, income, gender), as well as some characteristics of the actual trip (purpose, whether the user of the train had a Swiss annual season ticket, who paid for the travel costs, pieces of luggage being carried). Since both origin and destination were also collected, the dataset was completed with revealed preference data. In this thesis, however, we work with the stated preference data only. Further details about the survey design, data collection, and descriptive statistics can be consulted in Bierlaire et al. (2001). Our utility specification mimics the indirect utility considered for the stated preference responses in Bierlaire et al. (2001).

5.2 Results

5.2.1 Point Estimates

The MCMC estimator for the multinomial probit model with full covariance matrix and non-informative priors was run using 1,000, 10,000, 50,000, and 500,000 iterations separately. In addition, the maximum simulated likelihood estimator (MSLE) of the probit model was found using the GHK recursive transformation of the choice probabilities. Point estimates and standard errors are presented in Table 5.2.1. Note that Bayes point estimates and standard errors are given by the mean and standard

Table 5.2.1: Point estimates and standard errors for the travel mode choice case study

	MSLE		**1,000**		**10,000**		**50,000**		**50,0000**	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
ASC Swissmetro	0.1394	0.0360	0.2533	0.1731	0.1433	0.0757	0.1378	0.0435	0.1357	0.0344
ASC Car	-0.2714	0.0573	-0.1017	0.2554	-0.2664	0.1152	-0.2735	0.0682	-0.2769	0.0554
Travel Time [min]	-0.0018	0.0003	-0.0021	0.0004	-0.0018	0.0004	-0.0018	0.0003	-0.0018	0.0003
Travel Cost [CHF]	-0.0002	0.0000	-0.0002	0.0001	-0.0002	0.0001	-0.0002	0.0000	-0.0002	0.0000
Headway [min]	-0.0012	0.0002	-0.0017	0.0007	-0.0012	0.0004	-0.0012	0.0002	-0.0011	0.0002
Annual ticket	2.5553	0.1600	2.7557	0.4333	2.5368	0.2364	2.5439	0.1617	2.5401	0.1572
Age	0.0261	0.0057	0.0277	0.0116	0.0253	0.0069	0.0258	0.0056	0.0258	0.0056
Luggage	-0.3056	0.0335	-0.2881	0.0600	-0.3026	0.0372	-0.3056	0.0344	-0.3056	0.0336
Seats	-0.0799	0.0166	-0.0509	0.0978	-0.0741	0.0358	-0.0787	0.0215	-0.0792	0.0166
$\tilde{\Sigma}_{22}$	0.1182	0.0388	0.3631	0.4303	0.1351	0.1668	0.1195	0.0770	0.1150	0.0361
$\tilde{\Sigma}_{23}$	-0.2151	0.0471	-0.1079	0.2308	-0.2016	0.0869	-0.2157	0.0574	-0.2179	0.0454
$\tilde{\Sigma}_{33}$	1.8819	0.0388	1.6369	0.4303	1.8649	0.1668	1.8805	0.0770	1.8850	0.0361

deviation of the posterior, respectively. In the case of the nuisance parameters, the point estimates correspond to the covariance matrix in differences with respect to the first alternative.

Since non-informative priors were used, the posterior distribution of the parameters is fully determined by the likelihood function. This is a direct result of Bayes theorem, in which the posterior is proportional to the prior times the likelihood function. If the priors have a very low precision, then $p(\theta \mid y, X) \propto \ell(\theta \mid y, X)$, where $\ell(\theta \mid y, X)$ represents the conditional likelihood function. In fact, even with informative priors, the Bayes and maximum likelihood estimators coincide asymptotically. Since the asymptotic properties of the estimators are the same, if maximum likelihood estimates are available one can define the frequentist results as a benchmark to test for convergence. In this case, already with 10,000 iterations the Bayes point estimates are very close to the MSLE. However, the standard deviations of the posterior are much larger, an indication of the Bayes estimates not being fully efficient with this short run. With 50,000 iterations, the MSLE and Bayes point estimates are practically undistinguishable. The standard deviations of the posterior of the parameters of interest are slightly larger, but not by much. However, some of the standard deviations of the nuisance parameters almost double the standard errors obtained for the MSLE. Finally, using a long run of 500,000 both the point estimates and standard errors of the Bayes estimator and MSLE practically coincide. Note that the Bayes estimator exhibit standard errors that are, in general, somewhat lower than the MSLE standard errors. This may be due to the frequentist estimator being based on maximizing a simulated maximum likelihood instead of the actual likelihood function. In terms of statistical performance, the 95% credible intervals of all parameters contain the MSLE results, even for the very short run of just 1,000 iterations. However, in this case the intervals are wide. In fact, 5 of the 95% credible intervals also contain zero for the 1,000 iteration case.

Although the comparison with the MLE provides useful insights about the properties of the posterior distribution, the analysis of convergence should be based on the Bayesian results alone. The multinomial probit Gibbs sampler was rerun three different times -- with different starting values -- to perform the following series of standard convergence tests.

5.2.2 Visual Inspection

The first test for assessing convergence is visual inspection of traceplots, which display the sampled draws at every iteration for every parameter. Figure 5.2.2.1 summarizes the resulting traceplots for selected parameters and for the two extreme cases of 1,000 and 500,000 iterations of the Gibbs sampler.

The traceplots of the parameters show good behavior after 10,000 iterations, when the mean stabilizes and there are no indications of dependence on starting values. However, one needs to check the results of the formal tests to have betters insights about convergence of the chains.

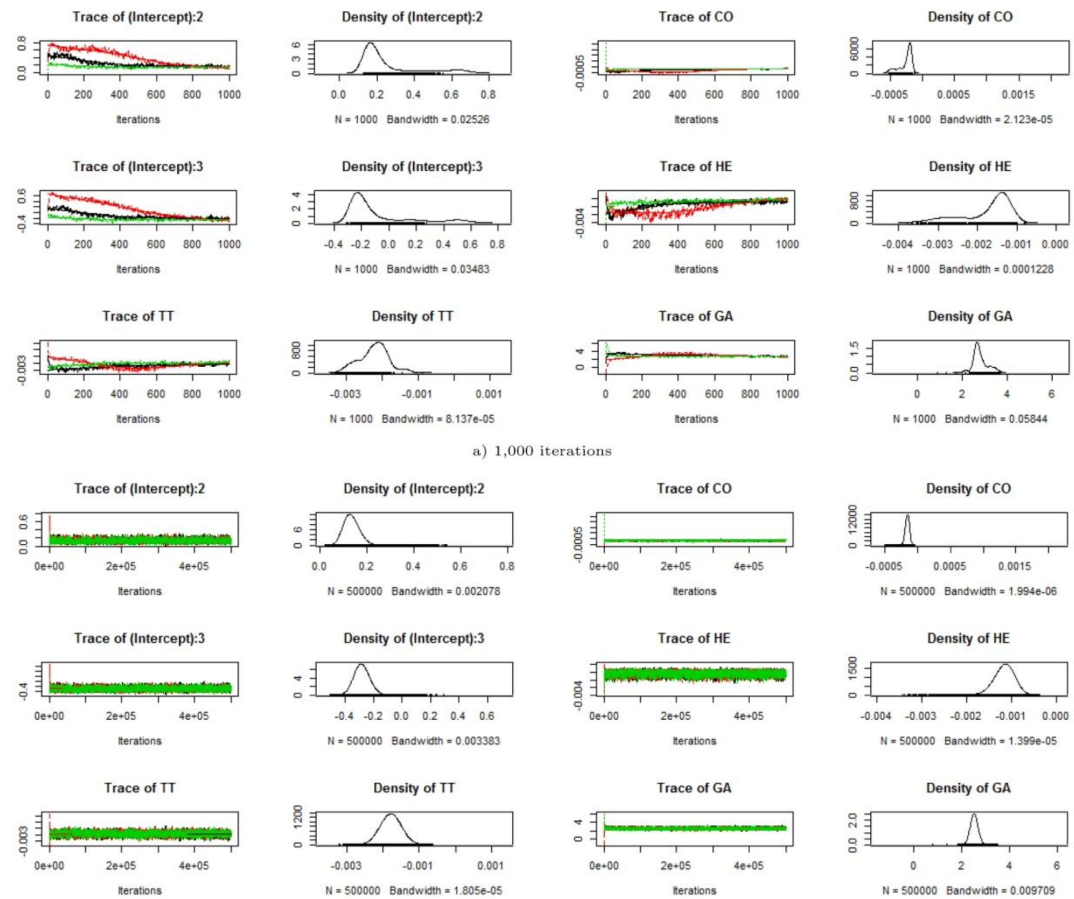
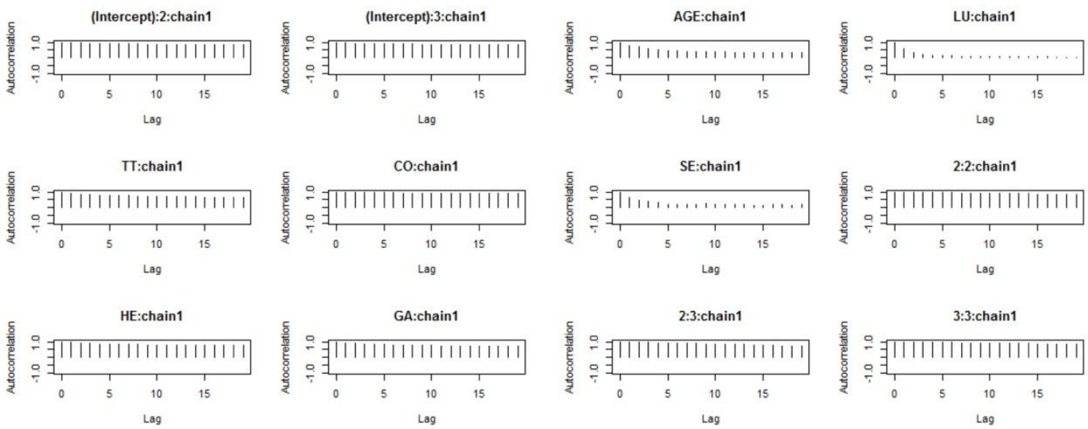


Figure 5.2.1: Traceplots and nonparametric density estimates of selected parameters

It is generally argued that traceplots mask the potential correlation among the different draws that are generated for estimation. Thus, a second common visual-inspection test is the analysis of autocorrelation, which usually complements the analysis of traceplots in applied work.

Figure 5.2.2.2 shows that most of the parameters have autocorrelations above 0.7 even when considering long lags. One of the problems of autocorrelation tests is that MCMC samples are not iid simply because they are built as a Markov chain. Additionally, as discussed above, even if correlations are high the ergodic theorem of

Markov chains will ensure convergence. Thus, high autocorrelations mean that it is safer to run a longer chain rather than being an indicator of poor mixing.



a) 1,000 iterations

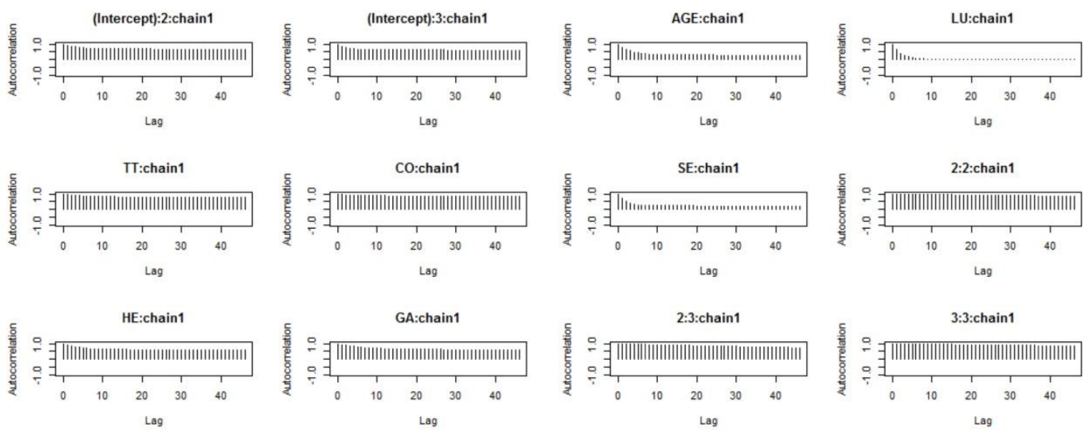


Figure 5.2.2: Autocorrelation for selected parameters

5.2.3 Geweke test of nonstationarity

The Geweke test compares the mean of two independent portions of the chain by calculating a Z-score statistic. The null hypothesis is equality of the means, and the test statistic is simply calculated as the difference of the sample means divided by the standard error (adjusted by autocorrelation). The asymptotic distribution of the test is a standard normal distribution. Table 5.2.2 presents the Z-scores for the first chains of the four runs considered.

For the short runs the Z-scores are very high, providing no evidence in favor of the null hypothesis. In the case of 50,000 iterations, seven of the parameters exhibit equal means (at the 95% confidence level). The longest run (500,000 iterations) passes the test of convergence for all parameters. Note that from the visual inspection test, the means seem equal starting at 10,000 iterations. The high Z-scores are due in this case to the high autocorrelations.

Table 5.2.2: Z-scores for the Geweke test

	MCMC iterations			
	1,000	10,000	50,000	500,000
ASC Swissmetro	30.149	21.248	2.188	1.363*
ASC Car	35.802	32.373	2.311	1.371*
Travel Time [min]	-5.264	13.493	-1.644*	-1.272*
Travel Cost [CHF]	-21.871	-2.495	-1.903*	-1.336*
Headway [min]	-12.246	-5.069	-1.612*	-1.127*
Annual ticket	5.897	-8.367	2.060	1.416*
Age	3.868	-12.848	1.702*	1.111*
Luggage	2.601	26.197	1.457*	0.376*
Seats	5.412	31.999	-0.566*	-0.969*
$\tilde{\Sigma}_{22}$	23.654	86.189	1.411*	1.209*
$\tilde{\Sigma}_{23}$	4.670	50.129	2.408	1.003*
$\tilde{\Sigma}_{33}$	-23.654	-86.189	-1.411*	-1.209*

* Cannot reject equality of the means at the 95% confidence level.

5.2.4 Raftery-Lewis diagnostic

We ran the Raftery-Lewis diagnostic for a pilot sampler of minimum length, given that we wanted to measure the 2.5% quantile of the posterior distribution with an acceptable tolerance of 0.005 and a probability of 0.95 of obtaining an estimate of the desired quantile within the preset accuracy. Results of the diagnostic are presented in Table 5.2.3.

The lower bound corresponds to the length of the pilot sampler, and is interpreted as the total number of iterations of an iid sampler. The burn-in iterations is the number of initial draws that should be erased before summarizing the posterior. The estimates of the burn-in iterations seem low, especially if one looks at the effect of differing starting values in Figure 5.2.2.1. High dependence factors come at no surprise given the high autocorrelations detected. The dependence factors amplify the lower bounds to derive an estimate of the total length required. If we take the maximum of the total length, the Raftery-Lewis diagnostic suggests using about 500,000 iterations, i.e. the longest run we have been analyzing. One of the problems with the diagnostic is that it overestimates the effect of autocorrelation and the estimates for total length are conservative.

Table 5.2.3: Raftery-Lewis diagnostic

	Burn-in	Total Length	Lower Bound	Dependence factor
ASC Swissmetro	27	29,832	3,746	7.96
ASC Car	30	39,045	3,746	10.40
Travel Time [min]	80	82,760	3,746	22.10
Travel Cost [CHF]	376	393,432	3,746	105.00
Headway [min]	270	317,475	3,746	84.80
Annual ticket	44	45,044	3,746	12.00
Age	16	17,204	3,746	4.59
Luggage	6	7,130	3,746	1.90
Seats	12	12,556	3,746	3.35
$\tilde{\Sigma}_{22}$	312	332,196	3,746	88.70
$\tilde{\Sigma}_{23}$	125	133,620	3,746	35.70
$\tilde{\Sigma}_{33}$	848	473,003	3,746	126.00

5.2.5 Gelman-Rubin diagnostic

The idea behind the Gelman-Rubin diagnostic is very intuitive. Different chains are built using over-dispersed starting values. Convergence is achieved when the outputs of the chains cannot be distinguished. An indicator of undistinguishable chains is given by potential scale reduction factors. Empirically, the 95% credible interval of the potential scale reduction factors is calculated to account for the uncertainty introduced by chains with finite lengths. Since lack of convergence will produce wide credible intervals, rather than just looking at the point estimates, the diagnostic looks at the 97.5% quantile (the upper bound of the 95% central credible interval). Table 5.2.4 presents the point estimates and upper bound of the credible interval of the Gelman-Rubin potential scale reduction factors.

Table 5.2.4: Gelman-Rubin potential scale reduction factors

	1,000		**10,000**		**50,000**		**500,000**	
	Est.	Upper C.I.	Est.	Upper C.I.	Est.	Upper C.I.	Est.	Upper C.I.
ASC Swissmetro	1.50	3.12	1.08	1.24	1.04	1.13	1.01	1.03
ASC Car	1.52	3.27	1.09	1.26	1.03	1.11	1.01	1.02
Travel Time	1.51	2.90	1.10	1.26	1.05	1.17	1.01	1.03
Travel Cost	1.63	4.89	1.13	1.36	1.05	1.16	1.01	1.03
Headway	1.32	2.19	1.07	1.16	1.03	1.11	1.01	1.02
Annual ticket	1.55	2.92	1.08	1.22	1.04	1.13	1.01	1.02
Age	1.03	1.08	1.03	1.09	1.01	1.05	1.00	1.01
Luggage	1.01	1.03	1.00	1.00	1.00	1.00	1.00	1.00
Seats	1.02	1.06	1.03	1.08	1.01	1.03	1.00	1.01
$\tilde{\Sigma}_{22}$	1.48	3.45	1.08	1.20	1.05	1.15	1.01	1.03
$\tilde{\Sigma}_{23}$	1.38	2.02	1.08	1.23	1.00	1.00	1.00	1.01
$\tilde{\Sigma}_{33}$	1.48	4.27	1.12	1.34	1.04	1.14	1.01	1.02

The 97.5% quantile of the Gelman-Rubin potential scale reduction factor of almost all parameters at 1,000 iterations exhibit values that show that the chain needs a longer run. In practice it is common to use a threshold of 1.2 for the upper bound of the credible interval. Thus, eight out of the twelve parameters fail to pass the convergence diagnostic at 10,000 iterations. With 50,000 iterations all parameters pass and the conclusion is that convergence has been achieved. The result is even stronger for the case of 500,000 iterations, where all upper bounds are well below 1.1 (another common threshold for the test.)

CHAPTER 6

DISCUSSION, CONCLUSIONS, AND FUTURE WORK

Convergence of MCMC draws to the stationary posterior distribution of interest is ensured under very general conditions. However, practical applications require testing for convergence after a finite number of repetitions of the sampler to make sure that inference is correct. A series of convergence diagnostics have been proposed in the literature, but there are only a few examples of application of these diagnostics to discrete choice estimators. These previous tests have emphasized problems, such as autocorrelation, that are no longer an issue. In addition, there is a general lack of convergence analysis of discrete choice models in willingness-to-pay space.

In this thesis we have analyzed the behavior of the four most common diagnostic tests for convergence, namely visual inspection, the Geweke test of nonstationarity, the Raftery-Lewis diagnostic, and the Gelman-Rubin test, not only in preference but also in WTP space. These tests were applied to a stated vehicle preference study that, with four alternatives, represents an average discrete choice experiment. When working with over-dispersed starting points, the influence of where the chain starts vanishes at 10-100 iterations, depending on the parameter. Total lengths of the chain beyond 3,746 iterations confirm that there are no trends in the traceplots and that there is good mixing in general, with quick drop in the autocorrelations and no need to thin the chains. As discussed in the thesis, some years ago diagnosing convergence was focused on breaking down the dependence among successive samples from the posterior. There was an excessive emphasis on thinning out the chain that is somewhat pervasive in empirical applications. Furthermore, thinning Markov chains to break down autocorrelation is not only computationally inefficient, but is also less precise. In addition, the ergodic theorem states that Monte Carlo integration works for chains that are correlated as long as these chains are irreducible, recurrent, and aperiodic.

From the visual inspection tests one can observe good behavior of the traceplots in preference space even at relatively short runs (3,746 iterations). However, for WTP space only chains from the probit Gibbs sampler show good visual convergence. In addition, because the stationary posterior distribution is asymptotically normal and actually coincides with the asymptotic distribution of the maximum likelihood estimator, we also compared the results of the Markov chain derived from MCMC estimators with the frequentist's estimates. Whereas the Bayes estimators provide information about the whole posterior, inference is based on the moments of this distribution. In effect, the Bayesian answer to the point estimation problem is the posterior mean. As a measure of uncertainty about the determination of the posterior mean, the posterior variance (or standard deviation) is calculated. The posterior mean coincides asymptotically with the true parameters, just as the frequentist estimator does. The posterior standard deviation can be interpreted as the standard error of the sampling distribution of the parameters. When we looked at the posterior means and compared them with the MSLE point estimates, we observed that estimates are closer even for longer iterations.

In terms of more formal convergence tests, the Geweke test of nonstationarity and the Gelman-Rubin test show good convergence for the results in preference space. Almost all the Z-scores and Gelman-Rubin potential scale reduction factors in WTP space could not pass the test, while post-processed WTP performs well in these two tests. The latter results contradict the conclusion from visual inspection. The reason for this contradiction is the existence of outliers, which introduce distortions in the the shape of the posterior chains. The frequent of outliers' appearance is not high actually (no more than 10 for 20,000 iterations), so with increase of length of chains, the influence of outliers decreases, contributing to pass of tests. Additionally, the stableness of point estimates with different iteration numbers support the conclusion of statistical tests on another aspect.

In sum, both the Geweke and Gelman-Rubin tests not only provide the same answer, but their results coincide with the behavior of frequentist estimates. The Raftery-Lewis diagnostic is another alternative, however, the results of this test are rather conservative. Still, 20,000 iterations are suggested for the multinomial logit and 50,000 for the mutinomial probit model. These numbers are below what is considered in practice (with 100,000-200,000 iterations being common). Finally, convergence diagnostics do not prove that convergence has been achieved, but the tests act as good tools to detect clear convergence problems, such as poor mixing or coverage of the parameter space.

In terms of further research, it would be interesting to analyze convergence in terms of the performance of the predictive posteriors of the choice probabilities and market shares, which are the relevant outputs for policy-oriented analysis with discrete choice. For instance, according to the comparison between the Bayes estimator and MSLE, direct calculation in each iterations appears to be accurate and efficient. The impact of the efficiency loss on the interval estimation of the market shares would be an interesting metric to study.

REFERENCES

Albert, J. H. and Chib, S. 1993 “Bayesian analysis of binary and polychotomous response data.” *Journal of the American Statistical Association*, 88 (422), 669–679.

Bierlaire, M., Axhausen, K. W. and Abay, G. 2001 “The acceptance of modal innovation : The case of Swissmetro.” In *Proceedings of the 1st Swiss Transport Research Conference*, Monte Verita, Ascona, Switzerland.

Bierlaire, M., Bolduc, D. and McFadden, D. 2008 “The estimation of generalized extreme value models from choice-based samples.” *Transportation Research Part B: Methodological*, 42(4), 381–394.

Bolduc, D., Boucher, N., and Daziano, R. A. 2008 “Hybrid choice modeling of new technologies for car choice in Canada.” *Transportation Research Record: Journal of the Transportation Research Board*, 2082 (1), 63–71.

Bolduc, D., Khalaf, L. and Ylou, C. 2010 “Identification robust confidence set methods for inference on parameter ratios with application to discrete choice models.” *Journal of Econometrics*, 157 (2), 317–327.

Brooks, S. P. and Roberts, G. O. 1997 “Assessing convergence of Markov chain Monte Carlo algorithms.” *Statistics and Computing*, 8 (4), 319–335.

Brownstone, D. 2001 “Discrete choice modeling for transportation.” *Travel Behaviour Research: The Leading Edge*, 97–124.

Chib, S., E. Greenberg, and Y. Chen, 1998 “MCMC methods for fitting and comparing multinomial response models.” *Econometrics 9802001*, EconWPA.

Cowles, M. K. and Carlin, B. P. 1996 “Markov chain Monte Carlo convergence diagnostics: a comparative review.” *Journal of the American Statistical Association*, 91 (434), 883–904.

Damien, P., Wakefield, J. and Walker, S. 1999 “Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables.” *Journal of the Royal Statistical Society*, 62(2), 331–344.

Daziano, R. A. and Bolduc, D. 2013 “Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model.” *Transportmetrica A: Transport Science*, 9 (1), 74–106.

Fosgerau, M. and Bierlaire, M. 2009 “Discrete choice models with multiplicatibe error terms.” *Transportation Research Part B: Methodological*, 43(5), 494–505.

Gelman, A. and Rubin, D. B. 1992 “Inference from iterative simulation using multiple sequences.” *Statistical Science*, 7 (4), 457–472.

Geweke, J. 1992 “Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments.” In *Bayesian Statistics (A. D. J.M. Bernardo,J.O. Berger and A. Smith, eds.)*, 169–193, Oxford University Press.

Godoy, G. and Ortzar, J. D. 2008 “On the estimation of mixed logit models.” *Transportation Research Trends*, 289–310.

Heidelberger, P. and Welch, P. 1983 “Simulation run length control in the presence of an initial transient.” *Operations Research*, 31, 1109–1144.

Hess, S., Bierlaire, M. and Polak, J. W. 2005 “Capturing Correlation and Taste Heterogeneity with Mixed GEV Models.” In *Applications of Simulation Methods in Environmental and Resource Economics (R. Scarpa, A. Alberini, and I. J. Bateman,* *eds.)*, Springer-Verlag, *The Economics of Non-Market Goods and Resources*, 6, 55–75.

Horne, M. 2003 "Incorporating preferences for personal urban transportation technologies into a hybrid energy-economy model." Masters degree thesis, Simon Fraser University, School of Resource and Environmental Management.

Lahiri, K. and Gao, J. 2002 "Bayesian analysis of nested logit model by Markov chain Monte Carlo." *Journal of Econometrics*, 111 (1), 103–133.

Link, W. A. and Eaton, M. J. 2012 "On thinning of chains in MCMC." *Methods in Ecology and Evolution*, 3 (1), 112–115.

MacEachern, S. N. and Berliner, L. M. 1994 "Subsampling the Gibbs sampler." *The American Statistician*, 48 (3), 188–190.

McCulloch, R. E., Polson, N. G. and Rossi, P. E. 2000 "A Bayesian analysis of the multinomial probit model with fully identified parameters." *Journal of the American Statistical Association*, 99 (1), 173–193.

McFadden, D. 2001 "Economic choices." *The American Economic Review*, 91 (3), 351–378.

Neal, R., 2003 "Slice sampling." *Annals of Statistics*, 3, 705–767.

Raftery, A. E. and Lewis, S. 1992 "How many iterations in the Gibbs sampler?" In *Bayesian Statistics (A. D. J.M. Bernardo,J.O. Berger and A. Smith*, *eds.)*, 763–773, Oxford University Press.

Rossi, P. E., Allenby, G. M. and McCulloch, R. 2005a *Bayesian Statistics and Marketing*. John Wiley and Sons, Chichester, West Sussex, UK.

Scott, S., 2003 “Data augmentation for the Bayesian analysis of multinomial logit models.” *Proceedings, American Statistical Association Section on Bayesian Statistical Science*, Alexandria, VA.

Train, K. 2009. *Discrete Choice Methods with Simulation*, 2nd edition, New York, NY:Cambridge University Press.

Train, K. and Weeks, M. 2005 “Discrete choice models in preference space and willingness-to-pay space.” In *Applications of Simulation Methods in Environmental and Resource Economics*, 1–16, Springer.

Wakefield, J., Gelfand, A. and Smith, A. 1991 “Efficient generation of random variates via the ratio-of-uniforms method.” *Statistics and Computing*, 1, 129–133.